Connecting Sentiment Analysis with Player Performance: Insights from Premier League Match Reports

William Wiik
732A81 - Text Mining
wilwi856@ad.liu.se

Abstract

This project investigates the relationship between sentiment in Premier League match reports and player ratings from FotMob. Using sentiment analysis with a pre-trained CardiffNLP Twitter-RoBERTa model, the study analyzes 158 match reports and correlates sentiment values assigned to players with their objective ratings. Sentences mentioning a single player were analyzed to assign sentiment scores, which were then aggregated to get an overall sentiment label for each player. The results reveal a clear trend: players with positive sentiment tend to have higher FotMob rating, while those with negative sentiment show significantly lower ratings.

1 Introduction

The objective of this project is to analyze player sentiment in football match reports and compare it with actual player ratings from FotMob and see if there is a correlation. This problem is interesting because it explores the relationship between tone in match reports and numerical performance ratings given to players, offering insight into how written sentiment aligns with statistical analyzes.

By solving this problem, we can gain a deeper understanding of how media coverage and player performance ratings correlate. This analysis can be valuable for sports analysts and coaches, helping them assess whether match reports reflect the actual performance of players as rated by statistical tools such as FotMob. The comparison also highlights potential mismatches between subjective descriptions of a player's performance and objective numerical ratings.

2 Theory

2.1 Sentiment Analysis

Sentiment analysis, is a natural language processing (NLP) task that aims to determine the sentiment or emotion expressed in a text (Wankhade et al., 2022).

2.2 CardiffNLP Twitter-RoBERTa

The sentiment analysis in this project uses the pre-trained model *cardiffnlp/twitter-roberta-base-sentiment*. The CardiffNLP model is based on the RoBERTa (Liu et al., 2019) architecture (Robustly Optimized BERT Pretraining Approach). RoBERTa improves the original BERT (Bidirectional Encoder Representations from Transformers) by

- Using larger mini batches for training.
- Removing the Next Sentence Prediction (NSP) task.
- Training on larger datasets with more robust preprocessing.

The model was trained on around 58 millions tweets and finetuned for sentiment analysis with the TweetEval benchmark (Barbieri et al., 2020). The model has three labels, negative, neutral and postive.

This project uses this model because football match reports are often written in an informal style, with player mentions, descriptions, and emotions, similar to the social media text the model was finetuned on.

3 Data

The dataset consists of match reports and player performance statistics for 158 Premier League matches. The data was collected through web scraping and the FotMob API. Match reports were web scraped from the official Premier League website `www.premierleague.com/results`. Player performance data was obtained from the FotMob API `www.fotmob.com`. This dataset contains player rating for each player who participated in the match.

The match reports and player performance data for the 158 Premier League matches were combined into one dataset. An example observation is provided in Listing 1.

Listing 1: Example observation from dataset

```
$hteam: Manchester United
$ateam: Fulham
$starttime: 2024-08-16T19:00:00.000Z
$score: 1 0

$home_players:
Name                 Min.played    Rating
Andre Onana          90            8.7
Noussair Mazraoui    81            7.8
... (other players listed similarly)

$away_players:
Name                  Min.played   Rating
Bernd Leno            90           7.22
Kenny Tete            90           6.54
... (other players listed similarly)

$report:
Joshua Zirkzee scored a dramatic late
winner on his Manchester United debut...
```

Figure 1 presents a histogram illustrating the distribution of FotMob ratings.

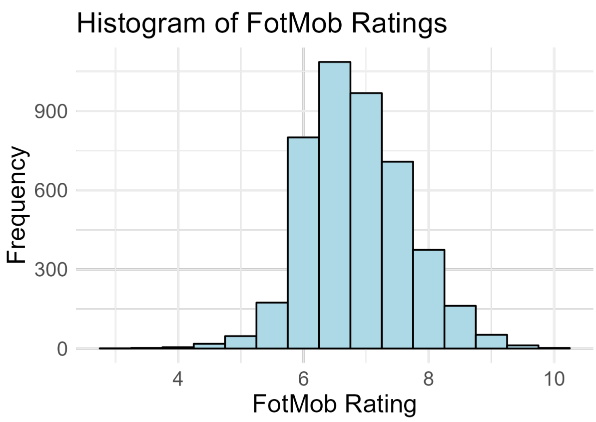

Figure 1: Histogram of FotMob Ratings

The histogram reveals that FotMob ratings typically fall within the range of 6 to 8.

3.1 Preprocessing

Each match report was split into individual sentences using a regular expression that detects sentence boundaries (periods, exclamation points, question marks). This step made it possible to analyze each sentence on its own, which is important in order to associating sentiment with the player mentioned in each sentence. Additionally, players with less than 10 minutes of playtime in a match were excluded, as a minimum of 10 minutes is required for a player to receive a player rating from FotMob.

4 Method

The process was implemented in Python, and the steps below are described to ensure reproducibility for others. A function was created with three arguments: `report`, `home_players` and `away_players`. The first step in the function involved processing the report and player data as outlined in the preprocessing section.

4.1 Analyze each sentence

Once the report and player data were preprocessed, the function iterated through each sentence in the report. Sentences mentioning two or more players, as well as those with no player mentions, were excluded from the analysis. This ensured that the analysis focused only on sentences containing a single player, allowing for a more accurate sentiment assignment. For sentences that met the criteria, the `the sentiment_model()` was used to determine the sentiment of the sentence, translating the sentiment label into a corresponding numerical score.

- LABEL_0 = negative = -1
- LABEL_1 = neutral = 0
- LABEL_2 = positive = 1

4.2 Aggregate sentiments for each player

After analyzing all sentences, the function aggregated the sentiment scores for each player. This was achieved by iterating through all players and their respective scores and calculating the average numerical score. The player's overall sentiment label was determined based on the average score:

- If the average score was greater than 0, the player was assigned the label **positive**
- If the average score was less than 0, the player was assigned the label **negative**.
- If the average score was exactly 0, the player was assigned the label **neutral**.

For example, if a player was mentioned in three sentences and received scores of -1, 0, and 1, their average score would be 0, resulting in a **neutral** sentiment label.

Once all the processing steps were completed, the results were stored in a dataframe containing three columns: `Player`, `Average Sentiment Label` and `FotMob Rating`.

5 Results

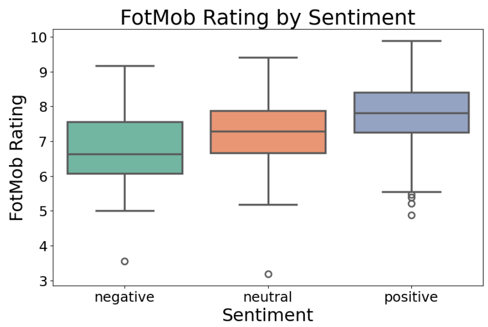

Figure 2: Boxplot of FotMob Ratings by Sentiment

The boxplot in Figure 2 visualizes the distribution of FotMob Ratings across three sentiment groups: Negative, Neutral, and Positive. Players with positive sentiment have the highest median and a narrow range of ratings. The neutral sentiment group shows a moderate median and a relatively compact spread of ratings. Players with negative sentiment have the lowest median and a wider range.

Table 1 presents a summary of FotMob Ratings categorized by sentiment. The Positive sentiment category includes 220 players, with a median rating of 7.81 and a mean rating of 7.77. The Neutral sentiment category has the largest number of players, with 313 individuals, showing a median rating of 7.28 and a mean rating of 7.27. The Negative sentiment category, with 79 players, has the lowest median rating of 6.63, and a mean rating of 6.75.

These results suggest that players with positive sentiments generally have higher FotMob ratings compared to those with neutral or negative sentiments. Furthermore, the findings indicate that players with negative sentiments tend to have the lowest ratings overall of the three groups.

Table 1: Summary of FotMob Ratings by Sentiment

Sentiment	n_players	Median	Mean
Positive	220	7.81	7.77
Neutral	313	7.28	7.27
Negative	79	6.63	6.75

6 Discussion

This study aimed to investigate the relationship between sentiment expressed in football match reports and player ratings from FotMob. The results suggest a strong positive correlation between positive sentiment and higher player ratings, while negative sentiments correspond to lower ratings. However, several challenges and limitations came during the analysis, which need to be acknowledged.

One significant limitation is the issue of player mentions in the reports. As the report only considered sentences where exactly one player was mentioned, many sentences had to be excluded from the analysis.

Furthermore, another issue came from the way players were referred to in the reports. In some cases, players were mentioned by their nicknames rather than their full names. The model was only able to identify players when they were mentioned by their real names, so it was impossible for the model to link sentiments to players referred to by their nicknames. This limitation may have led to a loss of information and potentially reduced the accuracy of sentiment assignment for certain players, which could have impacted the overall results.

A study by Rønningstad et al. (2024) investigates Entity-Level Sentiment Analysis (ELSA) and the challenges of determining sentiment towards specific entities in longer texts. It highlights the complexity of aggregating sentiments across multiple mentions and expressions of opinion. Similarly, this current project also faced challenges when attempting to assign sentiments to individual players within football match reports. Another issue both studies faced is the reliance on sentence-level sentiment analysis, which struggles with mixed sentiments in multi-entity sentences. To avoid confusion, this project excluded such sentences, but this reduced the usable dataset. The ELSA paper suggests more advanced methods like target-entity resolution to address this limitation.

7 Conclusion

This project demonstrates a significant correlation between sentiment expressed in Premier League match reports and players performance ratings. Positive sentiment aligns with higher FotMob ratings, while negative sentiment is associated with lower ratings, supporting the hypothesis that media sentiment reflects performance to some degree. By focusing on single-player mentions and utilizing the CardiffNLP Twitter-RoBERTa model, the analysis provides robust insights into how written narratives align with objective performance metrics.

References

Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *Preprint*, arXiv:2010.12421.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *Preprint*, arXiv:1907.11692.

Egil Rønningstad, Roman Klinger, Lilja Øvrelid, and Erik Velldal. 2024. Entity-level sentiment: More than the sum of its parts. *Preprint*, arXiv:2407.03916.

Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. 2022. A survey on sentiment analysis methods, applications, and challenges. *Preprint*, Springer:10.1007/s10462-022-10144-1.